

日 本 国 特 許 庁
JAPAN PATENT OFFICE

J1073 U.S. PTO

10/015352



12/12/01

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2000年12月14日

出 願 番 号

Application Number:

特願2000-380199

出 願 人

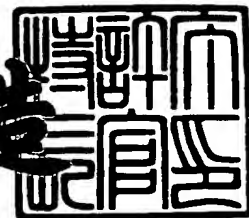
Applicant(s):

日本アイ・ビー・エム株式会社
株式会社大和銀行

2001年 6月 6日

特 許 庁 長 官
Commissioner,
Japan Patent Office

及 川 耕 造



出証番号 出証特2001-3051676

【書類名】 特許願

【整理番号】 JP9000349

【提出日】 平成12年12月14日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 7/00
G06F 17/40

【発明者】

【住所又は居所】 東京都港区六本木三丁目2番12号 日本アイ・ビー・エム株式会社内

【氏名】 鹿庭 義之

【発明者】

【住所又は居所】 大阪府大阪市中央区備後町2丁目2番1号 株式会社大和銀行内

【氏名】 山森 一頼

【特許出願人】

【識別番号】 592073101

【氏名又は名称】 日本アイ・ビー・エム株式会社

【特許出願人】

【識別番号】 393000054

【氏名又は名称】 株式会社大和銀行

【代理人】

【識別番号】 100104880

【弁理士】

【氏名又は名称】 古部 次郎

【選任した代理人】

【識別番号】 100100077

【弁理士】

【氏名又は名称】 大場 充

【手数料の表示】

【予納台帳番号】 081504

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 データ抽出方法、データ操作方法、債権情報抽出方法、データベースシステム、債権商品化処理装置、記憶媒体及びコンピュータプログラム

【特許請求の範囲】

【請求項 1】 コンピュータがデータベースに格納されている母体データ群から所定のデータの集合を抽出するデータ抽出方法において、

抽出すべき前記データの集合の要素である前記データの属性の構成比を設定するステップと、

前記データベースの前記母体データ群中に存在する前記属性を持つ対応データの量である抽出可能量を取得するステップと、

設定された前記構成比に基づいて前記属性ごとに前記データベースの前記母体データ群から抽出すべきデータ量である抽出データ量を計算するステップと、

前記構成比に基づいて算出された前記抽出データ量のうち、所定の前記属性における前記抽出データ量が当該属性を持つ前記対応データの抽出可能量を上回っている場合に、当該抽出データ量が前記抽出可能量以下となり、かつ前記構成比を一定の範囲で保持するように、当該抽出データ量を調整するステップとを含むことを特徴とするデータ抽出方法。

【請求項 2】 前記抽出データ量を調整する前記ステップは、

前記構成比が複数の構成条件ごとに設定されている場合に、当該複数の構成条件における前記属性の組合せごとに目標抽出量を計算するステップと、

前記構成条件の各々における前記構成比を変化させることなく、前記目標抽出量を調整することにより、前記抽出データ量を調整するステップとを含むことを特徴とする請求項 1 に記載のデータ抽出方法。

【請求項 3】 前記抽出データ量を調整する前記ステップは、

前記構成比が複数の構成条件ごとに設定されている場合に、当該複数の構成条件における前記属性の組合せごとに目標抽出量を計算する目標抽出量算出ステップと、

所定の二つの構成条件以外の構成条件における前記属性を特定の一つの属性に

固定し、当該二つの構成条件ごとに2種類ずつ選択された前記属性の組合せに対応する四つの前記目標抽出量を増減することにより、前記抽出データ量を調整する目標抽出量調整ステップと

を含むことを特徴とする請求項1に記載のデータ抽出方法。

【請求項4】 前記目標抽出量調整ステップは、

前記二つの構成条件に基づいて選択された前記四つの目標抽出量を、当該二つの構成条件における属性が両方とも異なる二つの目標抽出量からなる二つの組に分け、一方の組における二つの目標抽出量を所定量だけ減少させ、他方の組における二つの目標抽出量を当該所定量だけ増加させるステップを含むことを特徴とする請求項3に記載のデータ抽出方法。

【請求項5】 前記抽出データ量を調整する前記ステップは、

前記構成比を変えることなく、前記母体データ群から抽出するデータの全抽出量を減少させるステップをさらに含むことを特徴とする請求項1に記載のデータ抽出方法。

【請求項6】 データ処理装置が、記憶装置に記憶されているデータ群を読込んで操作を行うデータ操作方法において、

前記記憶装置から2種類の構成条件における属性ごとに定義されるデータ群を読み込み、

当該データ群のうち、構成条件ごとに2種類の属性を選んで、各属性の組合せに対応する4種類のデータ群を処理対象とし、

前記4種類のデータ群のうち、所定の一つのデータと、対応する二つの属性がいずれも当該所定の一つのデータと異なる他の一つのデータとを特定し、当該二つのデータの値を所定量だけ減少させ、

前記4種類のデータ群のうち、前記二つのデータ以外の二つのデータの値を前記所定量だけ増加させることにより、

前記構成条件の各々における構成比を変化させることなく、個々のデータの値を調整することを特徴とするデータ操作方法。

【請求項7】 3種類以上の構成条件における属性ごとに定義されるデータ群のうち、所定の2種類の構成条件以外の構成条件を特定の1属性に制限し、前

記 2 種類の種類構成条件における属性ごとに定義されるデータ群を特定することを特徴とする請求項 6 に記載のデータ操作方法。

【請求項 8】 データベースに蓄積された債権情報から、商品化の対象となる債権の集合を特定する債権情報を抽出する債権情報抽出方法において、

商品化する前記債権の集合に関して、当該集合の要素となる前記債権の属性の構成比を設定するステップと、

設定された前記構成比にしたがって、前記属性ごとに前記データベースから抽出すべき前記債権情報の数量である目標抽出量を計算するステップと、

算出された前記目標抽出量に基づいて、前記データベースから前記債権情報を抽出するステップと

を含むことを特徴とする債権情報抽出方法。

【請求項 9】 算出された前記目標抽出量のうち所定の前記属性における前記目標抽出量が、前記データベースに格納されている前記債権情報のうち当該属性を持つ債権の債権情報の数量を上回っている場合に、当該目標抽出量が前記データベースに格納されている当該属性に対応する債権情報の数量以下となり、かつ前記構成比を一定の範囲で保持するように、前記目標抽出量を調整するステップを含むことを特徴とする請求項 8 に記載の債権情報抽出方法。

【請求項 10】 母体データ群から所定の条件に応じてデータを抽出するデータベースシステムにおいて、

前記母体データ群を格納した抽出母体データベースと、

抽出条件としてデータの全抽出量と抽出データに対する複数の構成条件及び当該構成条件における属性の構成比とを含むデータ抽出要求を入力する抽出条件入力部と、

前記抽出条件入力部から入力した前記抽出条件に基づいて前記抽出母体データベースからデータを抽出するデータ処理部とを備え、

前記データ処理部は、

前記母体データ群中に存在する前記属性を持つ対応データの量である抽出可能量を前記抽出母体データベースに問い合わせ取得し、

前記抽出条件における全抽出量と前記属性の構成比とに基づいて前記属性ごと

に前記母体データ群から抽出すべきデータ量である抽出データ量を計算し、

前記構成比に基づいて算出された前記抽出データ量のうち、所定の前記属性における前記抽出データ量が当該属性を持つ前記対応データの前記抽出可能量を上回っている場合に、前記構成条件ごとの構成比を変化させることなく個々の前記属性ごとの前記抽出データ量を増減させることにより、当該抽出データ量が当該抽出可能量以下となるように当該抽出データ量を調整し、

前記調整された前記抽出データ量に基づいて前記母体データ群からデータを抽出することを特徴とするデータベースシステム。

【請求項 11】 前記データ処理部は、

前記複数の構成条件における前記属性の組合せごとに目標抽出量を計算し、

所定の二つの構成条件以外の構成条件における前記属性を特定の一つの属性に固定し、当該二つの構成条件ごとに2種類ずつ選択された前記属性の組合せに対応する四つの前記目標抽出量を増減することにより、前記抽出データ量を調整することを特徴とする請求項10に記載のデータベースシステム。

【請求項 12】 前記データ処理部は、

前記二つの構成条件に基づいて選択された前記四つの目標抽出量を、当該二つの構成条件における属性が両方とも異なる二つの目標抽出量からなる二つの組に分け、一方の組における二つの目標抽出量を所定量だけ減少させ、他方の組における二つの目標抽出量を当該所定量だけ増加させることを特徴とする請求項11に記載のデータベースシステム。

【請求項 13】 前記データ処理部は、

前記構成比に基づいて算出された前記抽出データ量のうち、所定の前記属性における前記抽出データ量が当該属性における前記抽出可能量を上回っている場合に、前記構成比を変えることなく、前記全抽出量を減少させることを特徴とする請求項10に記載のデータベースシステム。

【請求項 14】 データベースに蓄積された債権情報から、商品化の対象となる債権の集合を特定する債権情報を抽出する債権商品化処理装置において、

抽出条件として、商品化する前記債権の集合の要素となる前記債権の属性の構成比を入力する抽出条件入力部と、

設定された前記構成比にしたがって、前記属性ごとに前記データベースから抽出すべき前記債権情報の数量である目標抽出量を計算し、算出された前記目標抽出量に基づいて、前記データベースから前記債権情報を抽出するデータ処理部とを備えることを特徴とする債権商品化処理装置。

【請求項 1 5】 前記データ処理部は、

所定の前記属性における前記目標抽出量が前記データベースに格納されている前記債権情報のうち当該属性を持つ債権の債権情報の数量を上回っている場合に、当該目標抽出量が前記データベースに格納されている当該属性に対応する債権情報の数量以下となり、かつ前記構成比を一定の範囲で保持するように、前記目標抽出量を調整することを特徴とする請求項 1 4 に記載の債権商品化処理装置。

【請求項 1 6】 コンピュータに実行させるプログラムを当該コンピュータの入力手段が読取可能に記憶した記憶媒体において、

前記プログラムは、

抽出すべきデータの集合における当該データの属性の構成比を設定する処理と

母体データ群中に存在する前記属性を持つ対応データの量である抽出可能量を取得する処理と、

設定された前記構成比に基づいて前記属性ごとに前記母体データ群から抽出すべきデータ量である抽出データ量を計算する処理と、

前記構成比に基づいて算出された前記抽出データ量のうち、所定の前記属性における前記抽出データ量が当該属性を持つ前記対応データの前記抽出可能量を上回っている場合に、当該抽出データ量が前記抽出可能量以下となり、かつ前記構成比を一定の範囲で保持するように、当該抽出データ量を調整する処理とを前記コンピュータに実行させることを特徴とする記憶媒体。

【請求項 1 7】 コンピュータに実行させるプログラムを当該コンピュータの入力手段が読取可能に記憶した記憶媒体において、

前記プログラムは、

2 種類の構成条件における属性ごとに定義されるデータ群のうち、構成条件ごとに 2 種類の属性を選んで、各属性の組合せに対応する 4 種類のデータ群を処理

対象とし、前記 4 種類のデータ群のうち、所定の一つのデータと、対応する二つの属性がいずれも当該所定の一つのデータと異なる他の一つのデータとを特定し、当該二つのデータの値を所定量だけ減少させる処理と、

前記 4 種類のデータ群のうち、前記二つのデータ以外の二つのデータの値を前記所定量だけ増加させる処理とを前記コンピュータに実行させることを特徴とする記憶媒体。

【請求項 1 8】 抽出すべきデータの集合における当該データの属性の構成比を設定する処理と、

母体データ群中に存在する前記属性を持つ対応データの量である抽出可能量を取得する処理と、

設定された前記構成比に基づいて前記属性ごとに前記母体データ群から抽出すべきデータ量である抽出データ量を計算する処理と、

前記構成比に基づいて算出された前記抽出データ量のうち、所定の前記属性における前記抽出データ量が当該属性を持つ前記対応データの前記抽出可能量を上回っている場合に、当該抽出データ量が前記抽出可能量以下となり、かつ前記構成比を一定の範囲で保持するように、当該抽出データ量を調整する処理とをコンピュータに実行させるコンピュータプログラム。

【請求項 1 9】 2 種類の構成条件における属性ごとに定義されるデータ群のうち、構成条件ごとに 2 種類の属性を選んで、各属性の組合せに対応する 4 種類のデータ群を処理対象とし、前記 4 種類のデータ群のうち、所定の一つのデータと、対応する二つの属性がいずれも当該所定の一つのデータと異なる他の一つのデータとを特定し、当該二つのデータの値を所定量だけ減少させる処理と、

前記 4 種類のデータ群のうち、前記二つのデータ以外の二つのデータの値を前記所定量だけ増加させる処理とをコンピュータに実行させるコンピュータプログラム。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は、データベースから一定量のデータをランダムに抽出する手法として

、多次元の構成比を持つ結果データを選択する作為的抽出を行う方法に関する。

【 0 0 0 2 】

【従来の技術】

データベース中から一定量のデータをランダム抽出する代表的な手法として、データベース中のデータを乱数順に、一定量に至るまで無作為に取り出す無作為抽出がある。

データベース問合せシステムにおいては、取得しようとするデータに関して所定の項目が所定の値を取るというような条件を指定し、当該条件に合致するデータを抽出する場合がある。この場合においても、当該条件を満足するデータ群の中から、上記のような無作為抽出（ランダム抽出）によりデータ（結果データ）が取得される。

【 0 0 0 3 】

さらに、データベース中から複数の条件に対して特定の構成比を持つ結果データを抽出する方法としては、1. 抽出すべき一定量のデータ内容を少しずつ変化させながら目的の構成比のものを発見するまで繰り返す方法と、2. 目的の構成比となるように、対象条件を満足するデータの抽出量を条件ごとに予め決めておき、条件ごとに当該抽出量に相当する量のデータを抽出する方法とがある。

【 0 0 0 4 】

上記1の方法は、発見的手法であるので、最悪の場合、データの可能な組合せの全てを調べることになる。その計算コストは、 n 個のデータベースの場合、2の n 乗である。これは、 n が大きくなると指数関数的に計算コストが増えるため、計算機でも処理しきれなくなるNP問題（Non-Polynomial問題）という問題を引き起こす。したがって、数万件を超えるデータベースから結果データを抽出する場合の計算コストは天文学的数値となってしまう。

【 0 0 0 5 】

また上記2の方法は、例えば、条件1に関して値Aのデータと値Bのデータとの比が6：4となるような一定量1000のデータをランダム抽出したい場合を考える。この場合、条件1=Aのデータを600、条件1=Bのデータを400それぞれランダム抽出し、その後に得られたデータを足し合わせることで所

望の結果データが得られる。

結果データの構成比を定義付ける対象条件が2種類以上である場合（このような場合を2次元以上の構成比を持つという）は、各条件の組合せ（組合せ条件）を考え、各条件の構成比率の積を、当該組合せ条件ごとのデータの抽出量とする。例えば、条件1に関して値Aのデータと値Bのデータとの比が6：4であり、条件2に関して値Cのデータと値Dのデータとの比が比7：3となるようなデータを抽出したい場合を考える。

この場合、条件1と条件2の組合せとして、AC（条件1=A、条件2=C）、AD（条件1=A、条件2=D）、BC（条件1=B、条件2=C）、BD（条件1=B、条件2=D）の4種類が考えられる。

したがって、それぞれの条件の組合せにおける目標抽出量の比率を、ACのデータが42（ $=6 \times 7$ ）、ADのデータが18（ $=6 \times 3$ ）、BCのデータが28（ $=4 \times 7$ ）、BDのデータが12（ $=4 \times 3$ ）とすることができる。

【0006】

ところで、近年、金融機関などにおいて、資金調達などの目的で金銭債権の流動化が行われるようになってきている。かかる債権流動化のスキームは様々であるが、多くは、MBS（Mortgage Backed Securities）に代表されるような、多数の債権を集めてプール化し、それを背景とする証券等の商品化を行うものである。この場合、当該金融機関などが保有している債権の中から適当量の債権集合を抽出し、商品化するが、これらの商品の格付は、当該債権プールの統計的信用リスクに依存する。そのため、当該商品化の対象となる債権の集合に関して、予め構成条件を設定し、当該構成条件における債権の属性の構成比を設定して、保有している債権の集合の中から当該構成比を満足する債権の集合を抽出することができれば、目的とする格付をもつ商品を作成することができる。

このような商品化の対象となる債権の集合の抽出を行う場合、上述したような、データベース中から複数の条件に対して特定の構成比を持つ結果データを抽出する方法が必要となる。

【0007】

【発明が解決しようとする課題】

しかしながら、2次元以上の組合せ条件においては、当該条件間の相関関係によっては、各組合せ条件における抽出量を必ず満足できるとは限らない。反対に、各条件におけるデータの構成比率を満足する組合せ条件の抽出量が複数存在する場合もある。

例えば、上述した、条件1に関して値Aのデータと値Bのデータとの比が6:4であり、条件2に関して値Cのデータと値Dのデータとの比が比7:3となるようなデータを抽出したい場合、ACのデータが4、ADのデータが2、BCのデータが3、BDのデータが1の比率でも良いし、ACのデータが3、ADのデータが3、BCのデータが4、BDのデータが0の比率でも良い。前者の場合、条件1については

$$A : B = 6 (= 4 + 2) : 4 (= 3 + 1)$$

であり、条件2については

$$C : D = 7 (= 4 + 3) : 3 (= 2 + 1)$$

である。同様に後者の場合も、

条件1については

$$A : B = 6 (= 3 + 3) : 4 (= 4 + 0)$$

であり、条件2については

$$C : D = 7 (= 3 + 4) : 3 (= 3 + 0)$$

である。すなわち、いずれの場合も条件1に関する構成比率と条件2に関する構成比率とを満足している。

そして、組合せ条件において、各条件の構成比率の積を組合せ条件ごとのデータの抽出量とする場合、すなわち、

$$AC : AD : BC : BD = 4 \cdot 2 : 1 \cdot 8 : 2 \cdot 8 : 1 \cdot 2$$

の比率でデータの抽出ができない場合であっても、上記のような他の抽出量ならば抽出可能な場合もある。

【0008】

したがって、2次元以上の特定の構成比を持つ結果データを選択する作為抽出を正しく行うには、各条件に相関がある場合でも、可能な限り目標の構成比となるようにデータを抽出するため、組合せ条件ごとの比率（抽出量）を調整する手

段が必要となる。

【0009】

そこで、本発明は、データベースから一定量のデータを抽出する手法として、2次元以上の特定の構成比を持つ結果データを選択する作為的抽出を正しく行うことができるデータ抽出方法を提供することを目的とする。

【0010】

【課題を解決するための手段】

上記の目的を達成するため、本発明は、コンピュータがデータベースに格納されている母体データ群から所定のデータの集合を抽出するデータ抽出方法において、抽出すべきデータ集合の要素であるデータの属性の構成比を設定するステップと、前記データベースの母体データ群中に存在するこの属性を持つ対応データの量である抽出可能量を取得するステップと、設定された構成比に基づいてこの属性ごとに前記データベースの母体データ群から抽出すべきデータ量である抽出データ量を計算するステップと、この構成比に基づいて算出された抽出データ量のうち、所定の属性における抽出データ量がこの属性を持つ対応データの抽出可能量を上回っている場合に、この抽出データ量がこの抽出可能量以下となり、かつこの構成比を一定の範囲で保持するように、この抽出データ量を調整するステップとを含むことを特徴とする。

【0011】

ここで、この抽出データ量を調整するステップは、この構成比が複数の構成条件ごとに設定されている場合に、この複数の構成条件における属性の組合せごとに目標抽出量を計算するステップと、この構成条件の各々における構成比を変化させることなく、目標抽出量を調整することにより、抽出データ量を調整するステップとを含む構成とすることができる。

【0012】

また、この目標抽出量を調整するステップとして、所定の二つの構成条件以外の構成条件における属性を特定の一つの属性に固定し、この二つの構成条件ごとに2種類ずつ選択された属性の組合せに対応する四つの目標抽出量を増減する構成とすることができる。

これは、各構成条件を座標軸として想定される n 次元空間において、所定の二つの構成条件を座標軸とする 2 次元平面で切った断面において目標抽出量を操作する処理に相当する。したがって、構成条件がいくつ設定されていても、二つずつ構成条件を選択し、この二つの構成条件に関して処理を行うことにより、抽出データ量が抽出可能量以下となり、かつ設定された構成比を一定の範囲で（構成条件ごとの一次元構成比において）保持するように、抽出データ量を調整することができる。

【 0 0 1 3 】

さらに具体的には、この目標抽出量を調整するステップとして、この二つの構成条件に基づいて選択された四つの目標抽出量を、この二つの構成条件における属性が両方とも異なる二つの目標抽出量からなる二つの組に分け、一方の組における二つの目標抽出量を所定量だけ減少させ、他方の組における二つの目標抽出量をこの所定量だけ増加させる構成とすることができる。

【 0 0 1 4 】

さらにまた、この抽出データ量を調整するステップは、設定された構成比を変えることなく、この母体データ群から抽出するデータの全抽出量を減少させるステップをさらに含む構成とすることができる。

【 0 0 1 5 】

また本発明は、上記のデータ抽出方法を実現するデータベースシステムとして提供することができる。すなわち、このデータベースシステムは、母体データ群を格納した抽出母体データベースと、抽出条件としてデータの全抽出量と抽出データに対する複数の構成条件及びこの構成条件における属性の構成比とを含むデータ抽出要求を入力する抽出条件入力部と、この抽出条件入力部から入力した抽出条件に基づいてこの抽出母体データベースからデータを抽出するデータ処理部とを備え、このデータ処理部は、母体データ群中に存在するこの属性を持つ対応データの量である抽出可能量を抽出母体データベースに問い合わせ取得し、この抽出条件における全抽出量とこの属性の構成比とに基づいてこの属性ごとに母体データ群から抽出すべきデータ量である抽出データ量を計算し、この構成比に基づいて算出された抽出データ量のうち、所定の属性における抽出データ量がこ

の属性を持つ対応データの抽出可能量を上回っている場合に、構成条件ごとの構成比を変化させることなく個々の属性ごとの抽出データ量を増減させることにより、この抽出データ量がこの抽出可能量以下となるようにこの抽出データ量を調整し、この調整された抽出データ量に基づいて母体データ群からデータを抽出することを特徴とする。

さらに本発明は、コンピュータに上記の処理を実行させるコンピュータプログラムを作成し、このプログラムを記憶した記憶媒体として提供することができる。

【 0 0 1 6 】

また、本発明のデータ抽出方法は、データベースに蓄積された債権情報から、商品化の対象となる債権の集合を特定する債権情報を抽出する債権情報抽出方法として実現することができる。すなわち、この債権情報抽出方法は、商品化する債権の集合に関して、この集合の要素となる債権の属性の構成比を設定するステップと、設定された構成比にしたがって、この属性ごとにデータベースから抽出すべき債権情報の数量である目標抽出量を計算するステップと、算出された目標抽出量に基づいて、データベースからこの債権情報を抽出するステップとを含むことを特徴とする。

【 0 0 1 7 】

さらに本発明は、上記の債権情報抽出方法を実現する債権商品化処理装置として提供することができる。すなわち、この債権商品化処理装置は、抽出条件として、商品化する債権集合の要素となる債権の属性の構成比を入力する抽出条件入力部と、設定された構成比にしたがって、この属性ごとにデータベースから抽出すべき債権情報の数量である目標抽出量を計算し、算出された目標抽出量に基づいて、データベースから債権情報を抽出するデータ処理部とを備えることを特徴とする。

【 0 0 1 8 】

また、本発明は、データ処理装置が、記憶装置に記憶されているデータ群を読み込んで操作を行う、次のように構成されたデータ操作方法を提供することができる。すなわち、このデータ操作方法是、記憶装置から読み込んだ2種類の構成条

件における属性ごとに定義されるデータ群のうち、構成条件ごとに２種類の属性を選んで、各属性の組合せに対応する４種類のデータ群を処理対象とし、この４種類のデータ群のうち、所定の一つのデータと、対応する二つの属性がいずれもこの所定の一つのデータと異なる他の一つのデータとを特定し、この二つのデータの値を所定量だけ減少させ、この４種類のデータ群のうち、他の二つのデータの値を同じ所定量だけ増加させることにより、この構成条件の各々における構成比を変化させることなく、個々のデータの値を調整することを特徴とする。

【 0 0 1 9 】

このデータ操作方法は、構成条件が３種類以上設定されている場合、この３種類以上の構成条件における属性ごとに定義されるデータ群のうち、所定の２種類の構成条件以外の構成条件を特定の１属性に制限し、この２種類の構成条件における属性ごとに定義されるデータ群を特定することができる。

【 0 0 2 0 】

このデータ操作方法においても、コンピュータに上記のデータ操作を実行させるコンピュータプログラムを作成し、このコンピュータプログラムを記憶した記憶媒体として提供することができる。

【 0 0 2 1 】

【発明の実施の形態】

以下、添付図面に示す実施の形態に基づいて、この発明を詳細に説明する。

図１は、本実施の形態におけるデータベースシステムの全体構成を説明する図である。

図１を参照すると、本実施の形態のデータベースシステムは、データベースから所望のデータを抽出するための抽出条件を入力する抽出条件入力部１０と、データ抽出や抽出条件の加工などの処理を行うデータ処理部２０と、データ処理部２０における処理の経過や処理に用いる情報を記憶した条件別状況記憶部３０、属性値組合せ別状況記憶部４０及び乱数順キー記憶部５０と、抽出母体であるデータ群（以下、母体データ群と称す）を格納した抽出母体データベース６０と、抽出母体のデータ群から抽出されたデータ群（以下、結果データ群と称す）を格納する抽出結果データベース７０と、抽出母体データベース６０及び抽出結果デ

データベース70を管理するデータベース管理部80と、抽出結果データベース70に格納されている結果データ群を出力する出力部90とを備える。

【0022】

図1において、抽出条件入力部10は、例えばキーボードやマウスなどの入力デバイスと入力画面を表示するディスプレイ装置と入出力インタフェースとで実現され、母体データ群から結果データ群を得るための抽出条件を入力する。また、ネットワークを介して外部装置からの入力を受け付けるようにしても良いし、SQL問い合わせを入力するための対話式の入力手段などを用いても良い。

【0023】

データ処理部20は、例えばプログラム制御されたCPUとRAMその他のメモリとで実現される。本実施の形態におけるデータ処理部20は、抽出条件入力部10により入力された抽出条件と抽出母体データベース60に格納された母体データ群のデータ構成とに基づいて抽出するデータ量を決定したり、決定したデータ量に基づいて母体データ群からデータを抽出したり、外部からのデータ取得要求に応じて抽出結果データベース70に蓄積されている結果データ群を読み出して出力部90を介して出力したりする。これらの処理の具体的な内容については後述する。

【0024】

条件別状況記憶部30は、例えば半導体メモリや磁気記憶装置にて実現され、抽出条件入力部10により入力された抽出条件ごとに抽出しようとするデータ量（以下、目標抽出量と称す）や母体データ群から抽出可能なデータ量（以下、抽出可能量と称す）を対応付けた対応データを格納する。かかる対応データは、データ処理部20にて作成され、条件別状況記憶部30に格納される。

【0025】

属性値組合せ別状況記憶部40は、例えば半導体メモリや磁気記憶装置にて実現され、抽出条件入力部10により入力された抽出条件の組合せごとに目標抽出量や抽出可能量を対応付けた対応データを格納する。かかる対応データは、データ処理部20にて作成され、条件別状況記憶部30に格納される。

【0026】

乱数順キー記憶部 5 0 は、例えば半導体メモリや磁気記憶装置にて実現され、母体データ群からデータを抽出する際に、抽出された結果データ群を一時的に格納する。

【 0 0 2 7 】

抽出母体データベース 6 0 は、例えば半導体メモリや磁気記憶装置にて実現され、抽出対象である母体データ群を格納している。また、抽出結果データベース 7 0 は、例えば半導体メモリや磁気記憶装置にて実現され、母体データ群から抽出された結果データ群とを格納している。また、データベース管理部 8 0 は、例えばプログラム制御された CPU と RAM その他のメモリとで実現され、抽出母体データベース 6 0 及び抽出結果データベース 7 0 に対するアクセス（データの入出力）を管理する。また、出力部 9 0 は、ディスプレイ装置やプリンタ装置などの出力デバイスと入出力インタフェースとで実現され、抽出結果データベース 7 0 に格納された結果データ群を出力する。ネットワークを介して外部装置に結果データ群を出力するようにしても良い。

なお、CPU を制御してデータ処理部 2 0 やデータベース管理部 8 0 を実現するコンピュータプログラムは、CD-ROM やフロッピーディスクなどの記憶媒体に格納したり、ネットワークを介して伝送したりすることにより提供される。

【 0 0 2 8 】

次に、本実施の形態によるデータ抽出方法について説明する。

本実施の形態におけるデータの抽出は、抽出条件としてデータの構成条件と当該構成条件における構成比（当該構成条件において所定の値（以下、属性値と称す）を取るデータの結果データ群全体に対する割合）とを与え、この条件を満足する結果データ群を母体データ群から取得することにより行う。各条件を満足する個々のデータは、母体データ群からランダムに取得する。

図 2 乃至図 7 は、本実施の形態によるデータ抽出の動作を説明するフローチャートである。以下、これらのフローチャート及び図 8 乃至図 2 1 に示す図表を参照して本実施例によるデータ抽出方法を詳細に説明する。

【 0 0 2 9 】

初期動作として、抽出条件入力部 1 0 から抽出条件を入力して結果データの抽

出要求を行う。ここでは、図 8 に示す抽出条件が入力されたものとする。また、抽出母体データ群は 1 0 0 0 0 件のデータにて構成されるものとする。

図 8 を参照すると、全体抽出量として 1 0 0 0 件を抽出することとし、構成条件として A、B、C の三つの条件を指定している。また、構成条件 A に関して属性値 A 1 と属性値 A 2 のデータの構成比を 4 : 6、構成条件 B に関して属性値 B 1 と属性値 B 2 と属性値 B 3 のデータの構成比を 2 : 3 : 5、構成条件 C に関して属性値 C 1 と属性値 C 2 のデータの構成比を 5 : 5 とするように抽出条件を指定している（図 8 において、各構成条件における構成比はパーセント表示）。

【 0 0 3 0 】

結果データの抽出要求が入力されると、データ処理部 2 0 は、まず、上記構成条件及び各構成条件における属性値ごとに、抽出可能量を取得する（図 2、ステップ 2 0 1）。抽出可能量は、母体データ群に実際に存在する当該構成条件及び当該属性値に該当するデータの数を S Q L 問い合わせによりデータベース管理部 8 0 に問い合わせることにより取得することができる。

抽出可能量の取得動作は、全ての構成条件及び属性値に関して抽出可能量を取得するまで繰り返す（ステップ 2 0 2）。

【 0 0 3 1 】

次に、データ処理部 2 0 は、上記構成条件及び属性値ごとに、各属性値の構成比に対応する目標抽出量を算出する（ステップ 2 0 3）。目標抽出量は、全体抽出量と各構成条件における属性値の構成比との積により求める。例えば、構成条件 A においては、図 8 を参照すると属性値 A 1 のデータを 4 0 パーセント、属性値 A 2 のデータを 6 0 パーセント抽出するように指定されている。そこで、全体抽出量の 1 0 0 0 件を積算すれば、属性値 A 1 のデータにおける目標抽出量は 4 0 0 件、属性値 A 2 のデータにおける目標抽出量は 6 0 0 件となる。

図 9 は、以上のようにして得られた、構成条件及び属性値ごとの目標抽出量と抽出可能量との対応データの一例を示す図表である。図示の対応データは、条件別状況記憶部 3 0 に格納される。図 9 に示すように、本動作例では、母体データ群の 1 0 0 0 0 件のデータ中に、各構成条件における属性値ごとの抽出可能量は、構成条件 A で分類した場合、属性値 A 1、A 2 共に 5 0 0 0 件、構成条件 B で

分類した場合、属性値B 1が600件、属性値B 2が9000件、属性値B 3が400件、構成条件Cで分類した場合、属性値C 1が9700件、属性値C 2が300件、それぞれ存在するものとする。

【0032】

次に、データ処理部20は、図9に示す対応データにおいて各属性値における目標抽出量と抽出可能量とを比較し、目標抽出量の方が大きい項目が有るか調べる（ステップ204）。そして、入力された抽出条件における全ての構成条件及び属性値に関して目標抽出量が得られるまで、目標抽出量の算出及び得られた目標抽出量と抽出可能量との比較を繰り返す（ステップ205）。

【0033】

ステップ204において、目標抽出量の方が抽出可能量よりも大きい項目があった場合、そのままでは、抽出可能量のデータを全て抽出しても目標抽出量には達しない。したがって、抽出条件を満足するように結果データ群のデータ構成を抽出するためには、全ての項目における目標抽出量が抽出可能量以内に収まるように、全体抽出量を減少させて調整する。具体的には、次の式により、新たな全体抽出量を定義する（ステップ206）。

新たな全体抽出量＝以前の全体抽出量／目標抽出量＊抽出可能量

そして、得られた新たな全体抽出量に基づいて目標抽出量を再計算する。

図9に示した例では、構成条件Bにおける属性値B 3及び構成条件Cにおける属性値C 2において、目標抽出量の方が抽出可能量よりも大きくなっている。そこで、これらの項目で目標抽出量が抽出可能量以下となるように調整し、全体抽出量を600（＝1000／500＊300）件とする。

【0034】

このように、抽出条件である全体抽出量の設定から調整をやり直すことにより、各構成条件における属性の構成比を守った結果データ群を得ることができる。しかしながら、この調整によれば、各構成条件における属性の構成比に関しては、入力された抽出条件を満足しているが、全体抽出量に関しては抽出条件よりも

小さい値となる。そこで、各構成条件における属性の構成比が崩れてしまっても、サンプルとしての結果データ群の抽出量が重要であるような処理においては、ステップ204の判断及びステップ206における全体抽出量の調整を行う必要はない。

図10は、かかる調整を行った後の目標抽出量と抽出可能量との対応データを示す図表である。図10に示すように、属性値B3、C2のいずれも目標抽出量が抽出可能量以下となっている。この対応データは、図9の対応データに替えて条件別状況記憶部30に格納される。

【0035】

次に、データ処理部20は、各構成条件A、B、Cにおける各属性値の組合せ（以下、属性値組合せと称す）を求め、この属性値組合せごとのバランス比率と当該バランス比率に対応した目標抽出量と抽出可能量とを算出する（図3、ステップ207、208、209）。ここで、バランス比率とは、各属性値組合せにおける各属性値に対する構成比の積である。例えば、属性値組合せがA1、B1、C1の場合のバランス比率は、

$$40\% \times 20\% \times 50\% = 4\%$$

である。また、目標抽出量は、全体抽出量と各属性値組合せのバランス比率との積により算出される。また、抽出可能量は、母体データ群における該当データの実際の数とSQL問い合わせにてデータベース管理部80に問い合わせることにより取得することができる。

これらの情報の取得動作は、全ての構成条件及び属性値に関して取得するまで繰り返す（ステップ210）。

図11は、図10の対応データに基づいて算出された属性値組合せごとのバランス比率と目標抽出量と抽出可能量との対応データを示す図表である。図示の対応データは、属性値組合せ別状況記憶部40に格納される。図11において、全体抽出量は、条件別状況記憶部30に格納された対応データから取得することができる。

【0036】

次に、データ処理部20は、図11に示す対応データにおいて各属性値における目標抽出量と抽出可能量とを比較し、目標抽出量の方が大きい項目が有るか調べる（ステップ211）。全ての項目で目標抽出量が抽出可能量以下である場合は、母体データ群から実際にデータを抽出する図7のステップ235以降の処理に移行する（ステップ212）。

目標抽出量の方が抽出可能量よりも大きい項目があった場合、次に、図11に示したような属性値組合せごとの目標抽出量と抽出可能量との対応データを用いて目標抽出量の対角入替調整を行う。

【0037】

ここで、対角入替調整の基本概念について説明する。

図12に示すようなデータ群を考える。このデータ群は、条件 α と条件 β という2種類の構成条件に対応している（すなわち、2次元の構成比を持つ）。条件 α には $\alpha 1$ 、 $\alpha 2$ 、 $\alpha 3$ という3種類の属性値があり、条件 β には $\beta 1$ 、 $\beta 2$ という2種類の属性値がある。なお、図12において、図12（A）は対角入替調整前の状態を示し、図12（B）は対角入替調整後の状態を示す。

図12（A）のデータ群において、属性値組合せが $\alpha 1 \beta 1$ 、 $\alpha 2 \beta 1$ 、 $\alpha 1 \beta 2$ 、 $\alpha 2 \beta 2$ の四つのセルからなる長方形を考える。この長方形における所定の対角に位置する二つのセルの値に所定の値を加え、他の二つのセルの値から同じ値を減じた場合、図12（B）に示すように、各構成条件の1次元構成比は変わらない。このような操作を対角入替調整と称す。

【0038】

図12（A）、（B）を参照して具体的に説明する。

図12（A）において、属性値組合せ $\alpha 1 \beta 1$ のセルの値（＝16）とこのセルの対角に位置する属性値組合せ $\alpha 2 \beta 2$ のセルの値（＝18）とに値2を加え、他の二つのセルの値（12、24）から同じ値2を減じる。かかる対角入替調整を行っても、図12（B）に示すように、属性値 $\alpha 1$ であるデータの合計値（＝40）、属性値 $\alpha 2$ であるデータの合計値（＝30）、属性値 $\beta 1$ であるデータの合計値（＝40）、属性値 $\beta 2$ であるデータの合計値（＝60）は変化して

いない。

【 0 0 3 9 】

このような関係は、図 1 2 に示すような 2 次元の構成比を持つデータ群を対象とした場合、任意に想定された長方形における二つの対角に位置する四つのセルにおいて常に成り立つ。

また、2 次元以上の多次元の構成比を持つデータ群においても、当該データ群から任意の 2 次元の構成条件に着目することによって、対角入替調整を行うことが可能である。すなわち、処理対象とする 2 次元の構成条件以外の構成条件に関しては、適当な属性値に固定しておくことにより、当該 2 次元の構成条件の問題として考えることができる。これは、 n 種類の構成条件のそれぞれを座標軸とする n 次元空間を考え（空間内の各点が各属性値組合せに対応する）、当該 n 次元空間を任意の 2 次元平面（座標平面）で切り、当該 2 次元平面上で対角入替調整を行うことに相当する。

【 0 0 4 0 】

データ処理部 2 0 は、図 1 1 に示したような対応データに対して、上記の対角入替調整を行うことにより、各一次元構成比を変えることなく、目標抽出量の調整を行う。

図 1 3 は、図 1 1 に示した対応データを、各構成条件を座標軸とする n 次元（この場合は 3 次元）空間として認識したイメージを示す図表である。同図において、構成条件 A については図表を 2 段にして表現し、他の構成条件 B、C についてはそれぞれ図表の縦軸と横軸とを割り当てて表現している。

【 0 0 4 1 】

データ処理部 2 0 は、まず、図 1 3 に示すような対応データのうち、所定の 1 点（1 つのセル）を基準点とする（図 4、ステップ 2 1 3）。この基準点は、対角入替調整の対象であるセルの一つである。次に、構成条件のうちの所定の一つを 2 次元平面（座標平面）における一つめの軸（座標軸）とする（ステップ 2 1 4）。そして、当該一つめの軸における属性値のみが基準点と異なり、他の属性値は基準点と同一であるような 1 点（セル）を選択し、対角入替調整の対象であるセルとする（ステップ 2 1 5）。

次に、上記一つめの軸とは異なる構成条件のうちの所定の一つを 2 次元平面において当該一つめの軸と直交する二つめの軸とする（ステップ 2 1 6）。そして、当該二つめの軸における属性値のみが基準点と異なり、他の属性値は基準点と同一であるような 1 点（セル）を選択し、対角入替調整の対象であるセルとする（ステップ 2 1 7）。

最後に、一つめの軸における属性値がステップ 2 1 5 で選択した点と同一であり、二つめの軸における属性値がステップ 2 1 7 で選択した点と同一であり、他の属性値は基準点と同一であるような 1 点（セル）を対角入替調整の対象であるセルとする（ステップ 2 1 8）。

【 0 0 4 2 】

以上のようにして得られた四つのセルが、当該 2 次元平面において、上述した対角入替調整の対象である長方形の対角を構成する。なお、ここでは二つの軸が直交するとしているので、当該四つのセルが長方形をなすこととしたが、より一般的には平行四辺形をなすと考えることができる。

また、対角入替調整の対象となる四つのセルを選択する手順として、上記の動作例では、一つの基準点（セル）を特定し、構成条件及び属性値における基準点との関係を用いて他の三つのセルを選択した。しかしながら、この手法は一例に過ぎず、2 種類の構成条件を 2 軸とする 2 次元平面において所定の長方形（平行四辺形）の頂点をなす四つのセルを選択することができれば、どのような手法を用いてもかまわない。

【 0 0 4 3 】

次に、データ処理部 2 0 は、これらの四つのセルに対して値の操作を行う。まず、基準点において、目標抽出量が抽出可能量を上回っているかどうかを判断し（図 5、ステップ 2 1 9）、上回っていれば基準点の目標抽出量の調整方向（増加させるか減少させるか）を減少させる方向とする（ステップ 2 2 0）。調整量は目標抽出量から抽出可能量を減じた値（＝目標抽出量－抽出可能量）とする。

一方、基準点における目標抽出量が抽出可能量以下である場合、次に、基準点の目標抽出量の値がマイナスになっていないかを判断し（ステップ 2 2 1）、マイナスになっていたならば、基準点の目標抽出量の調整方向を増加させる方向と

する（ステップ 2 2 2）。調整量は調整可能な最大量、すなわち、基準点の対角に位置するセルにおける目標抽出量に加えたり、他の二つのセルにおける目標抽出量から減じたりできる最大量とする。なお、所定のセルにおける目標抽出量に関しては、当該目標抽出量が抽出可能量を超えない範囲で増加することが可能であり、また当該目標抽出量の値がマイナスにならない範囲で減少させることが可能である。

ここで、目標抽出量の値がマイナスになる場合とは、基準点における目標抽出量を抽出可能量以下に抑えるために調整した際に、この基準点の対角に位置するセルにおいて、目標抽出量を基準点と同じ値だけ減じたことによって目標抽出量が 0 を下回る場合に生じる。

【 0 0 4 4 】

また、ステップ 2 2 1 の判断において、基準点の目標抽出量が 0 よりも大きい、すなわち 0 と抽出可能量との間の値であるならば、当該基準点の調整可能量を算出する。調整可能量とは、目標抽出量と 0 との差および目標抽出量と抽出可能量との差である。そして、二つの調整可能量の大きい方を調整方向とし、当該調整可能量の 2 分の 1 の値を調整量とする（ステップ 2 2 3）。

なお、ステップ 2 2 0、2 2 2 で設定した調整量と、ステップ 2 2 3 で設定した調整方向及び調整量とは例示に過ぎず、目標抽出量を調整する目的に適した他の調整量を設定することも可能である。特にステップ 2 2 3 における調整方向及び調整量は、目標抽出量を、母体データ群に対する抽出可能量の割合に近づけるために設定されている。すなわち、この調整を行わなくても、目標抽出量を抽出可能量以下にするという対角入替調整の目的は損なわれない。そこで、結果データ群における各構成条件の属性値の構成比を母体データ群におけるデータの割合に近づける必要がない場合は、ステップ 2 2 3 における調整自体をスキップすることもできる。

【 0 0 4 5 】

ステップ 2 2 0、2 2 2、2 2 3 において基準点における目標抽出量の調整方向及び調整量を決定したならば、データ処理部 2 0 は、次に、決定された調整方向に応じて四つのセルにおける目標抽出量を調整する。すなわち、基準点におけ

る目標抽出量の調整方向が減少させる方向であれば、基準点とその対角に位置するセルにおける目標抽出量から調整量に相当する値を減じ、他の二つのセルにおける目標抽出量に調整量に相当する値を加える（ステップ 2 2 4、2 2 5）。

一方、基準点における目標抽出量の調整方向が増加させる方向であれば、基準点とその対角に位置するセルにおける目標抽出量から調整量に相当する値を加え、他の二つのセルにおける目標抽出量に調整量に相当する値を減ずる（ステップ 2 2 4、2 2 6）。

【 0 0 4 6 】

以上の動作を、図 1 3 を参照して具体的に説明する。ここでは、図 1 3 に示したデータ群のうち、構成条件 A の属性値を A 2 とする 2 次元平面（A 2 平面と称す）について考える。

まず、ステップ 2 1 3 において属性値組合せが A 2 B 1 C 2 であるセルを基準点とする。次に、ステップ 2 1 4 において構成条件 B を一つめの軸とし、ステップ 2 1 5 において構成条件 B の属性値が B 3 であるセル（すなわち属性値組合せが A 2 B 3 C 2 のセル）を対角入替調整の対象とする。さらに、ステップ 2 1 6 において構成条件 C を二つめの軸とし、ステップ 2 1 7 において構成条件 C の属性値が C 1 であるセル（すなわち属性値組合せが A 2 B 1 C 1 のセル）を対角入替調整の対象とする。次に、ステップ 2 1 8 において、構成条件 B の属性値が B 3 であり構成条件 C の属性値が C 1 であるセル（すなわち属性値組合せが A 2 B 3 C 1 のセル）を対角入替調整の対象とする。

図 1 3 の 2 本の矢印によって対応付けられた四つのセルが対角入替調整の対象であり、各矢印が指し示す二つのセルの組が、それぞれ対角に位置するセルである。

【 0 0 4 7 】

次に、基準点とした A 2 B 1 C 2 のセルについて目標抽出量を調べ、調整方向と調整量とを決定する。図 1 3 を参照すると、当該セルの目標抽出量が 3 6、抽出可能量が 0 であるから、ステップ 2 2 0 において調整方向を、目標抽出量を減少させる方向とし、調整量を、3 6（ $= 3 6 - 0$ ）とする。そして、調整方向が目標抽出量を減少させる方向であるので、ステップ 2 2 5 において、基準点とそ

の対角に位置するセルにおける目標抽出量から36を減じ、他の二つのセルにおける目標抽出量に36を加える。

【0048】

図14に、以上の操作を行った状態の対応データを示す。図13と図14とを比較すると、基準点としたA2B1C2のセルでは目標抽出量が36減少して0となり、抽出可能量と一致している。また、基準点の対角に位置するA2B3C1のセルにおいても目標抽出量が36減少して54となっている。また、A2B1C1のセルにおいては目標抽出量が36増加して72となり、A2B3C2のセルにおいては目標抽出量が36増加して126となっている。そして、いずれのセルにおいても目標抽出量が0以上で抽出可能量以下となっている。

【0049】

次に、データ処理部20は、基準点を含む長方形を変形し、また基準点を移動させながら、全てのセルの組合せに対して上記の対角入替調整を実行する。すなわち、まずステップ216で選択した二つめの軸について、当該二つめの軸における属性値のみが基準点と異なる未処理の点（セル）が他にあるかどうかを調べ、他にそのような点があれば、ステップ217に戻って対角入替調整の対象となるセルを改めて選択し、以降の処理を再帰的に繰り返す（図6、ステップ227）。

当該二つめの軸においてそのような未処理の点が無い場合、次に、当該二つめの軸以外に一つめの軸とは異なる未処理の構成条件があるかどうかを調べ、そのような未処理の構成条件があれば、ステップ216に戻って2次元平面における二つめの軸を改めて決定し、以降の処理を再帰的に繰り返す（ステップ228）。

そのような未処理の構成条件が無い場合、次に、ステップ214で選択した一つめの軸について、当該一つめの軸における属性値のみが基準点と異なる未処理の点（セル）が他にあるかどうかを調べ、他にそのような点があれば、ステップ215に戻って対角入替調整の対象となるセルを改めて選択し、以降の処理を再帰的に繰り返す（ステップ229）。

当該一つめの軸においてそのような未処理の基準点が無い場合、次に、一つめ

の軸とは異なる未処理の構成条件があるかどうかを調べ、そのような未処理の構成条件があれば、ステップ214に戻って2次元平面における一つめの軸を改めて決定し、以降の処理を再帰的に繰り返す（ステップ230）。

当該一つめの軸においてそのような未処理の点が無い場合、次に、未だ基準点としていない点（セル）があるかどうかを調べ、そのような未処理の点があれば、ステップ213に戻って基準点を改めて選択し、以降の処理を再帰的に繰り返す（ステップ231）。

【0050】

図14乃至図16を参照して、基準点や長方形を切り替えながら対角入替調整を行う例を具体的に説明する。

図14において、属性値組合せがA2B2C2であるセルを基準点とする。そして、構成条件Bと構成条件Cとを直交する座標軸とするA2平面において、属性値組合せがB2C1、B2C2（基準点）、B3C1、B3C2のセルで形成される長方形について対角入替調整を行う。図14の2本の矢印が指し示す二つのセルの組が、それぞれ対角に位置するセルである。

基準点としたA2B2C2のセルについて目標抽出量を調べると、当該セルの目標抽出量が54であり、抽出可能量が0であるから、調整方向を、目標抽出量を減少させる方向とし、調整量を、54（＝54－0）とする。そして、当該基準点とその対角に位置するセルにおける目標抽出量から54を減じ、他の二つのセルにおける目標抽出量に54を加える。

【0051】

図15は、以上の操作を行った状態の対応データを示す。図14と図15とを比較すると、基準点としたA2B2C2のセルでは目標抽出量が54減少して0となり、抽出可能量と一致している。また、基準点の対角に位置するA2B3C1のセルにおいても目標抽出量が54減少して0となっている。また、A2B2C1のセルにおいては目標抽出量が54増加して108となり、A2B3C2のセルにおいては目標抽出量が54増加して180となっている。そして、いずれのセルにおいても目標抽出量が0以上で抽出可能量以下となっている。

【0052】

次に、図15において、属性値組合せがA1B3C1であるセルを基準点とする。そして、構成条件Aと構成条件Bとを座標軸とするC1平面において、属性値組合せがA1B2、A1B3（基準点）、A2B2、A2B3のセルで形成される長方形について対角入替調整を行う。図15の2本の矢印が指し示す二つのセルの組が、それぞれ対角に位置するセルである。

基準点としたA1B3C1のセルについて目標抽出量を調べると、当該セルの目標抽出量が60であり、抽出可能量が0であるから、調整方向を、目標抽出量を減少させる方向とし、調整量を、60（＝60－0）とする。そして、当該基準点とその対角に位置するセルにおける目標抽出量から60を減じ、他の二つのセルにおける目標抽出量に60を加える。

【0053】

図16は、以上の操作を行った状態の対応データを示す。図15と図16とを比較すると、基準点としたA1B3C1のセルでは目標抽出量が60減少して0となり、抽出可能量と一致している。また、基準点の対角に位置するA2B2C1のセルにおいても目標抽出量が60減少して48となっている。また、A1B2C1のセルにおいては目標抽出量が60増加して96となり、A2B3C1のセルにおいては目標抽出量が60増加して60となっている。そして、いずれのセルにおいても目標抽出量が0以上で抽出可能量以下となっている。

【0054】

以上のようにして、属性値組合せごとの対応データの全ての組合せにおいて対角入替調整を行った後、データ処理部20は、対角入替調整を反映させた属性値組合せごとの対応データに基づいて、対角入替調整後の目標抽出量が抽出可能量を上回っている項目があるか調べる（ステップ232）。全ての項目で目標抽出量が抽出可能量以下である場合は、母体データ群から実際にデータを抽出する図7のステップ235以降の処理に移行する（ステップ233）。

【0055】

一方、目標抽出量が抽出可能量を上回っている項目がある場合、ステップ213に戻って対角入替調整を繰り返す。この繰り返しは予め定められた一定回数に達するまで行うこととする。そして、当該一定回数繰り返しても目標抽出量が抽

出可能量を上回っている項目が残っている場合（収束しない場合）、当該抽出条件によっては収束しないものとみなして、母体データ群から実際にデータを抽出する図7のステップ235以降の処理に移行する（ステップ234）。

【0056】

図17は、図11に示した属性値組合せごとの対応データに、図16までの対角入替調整を行った後の目標抽出量を加えた対応データを示す図表である。図17を参照すると、対角入替調整後の目標抽出量が抽出可能量を上回る項目が残っている（属性値組合せがA1B1C2やA1B2C2など）。したがって、対角入替調整を再度実行する。

図18は、対角入替調整を再度実行した状態の対応データを示す。また、図19は、図11に示した属性値組合せごとの対応データに、図18までの対角入替調整を行った後の目標抽出量を加えた対応データを示す図表である。

図19を参照すると、全ての項目において目標抽出量が抽出可能量以下となっている。したがって、この時点で対角入替調整を終了し、母体データ群から実際にデータを抽出する処理に移行する。

【0057】

なお、ステップ232、234の判断において、対角入替調整を一定回数以上行っても目標抽出量が抽出可能量を上回っている項目が残っている場合、抽出条件として入力された各構成条件における属性の構成比を優先させるような調整を行うこともできる。すなわち、ステップ206で行ったように、

新たな全体抽出量＝以前の全体抽出量／目標抽出量＊抽出可能量

という計算式により、新たな全体抽出量を算出し、これに基づいて目標抽出量を計算し直して、対角入替調整を再度実行する。このように、抽出条件である全体抽出量の設定から調整をやり直すことにより、この調整によって全体抽出量の値が0になってしまわない限り、各構成条件における属性の構成比を守った結果データ群を得ることができる。

【0058】

次に、データ処理部 20 は、抽出母体データベース 60 に格納されている母体データ群から、各属性値組合せに合致するデータを、目標抽出量分だけ抽出する。データの抽出は、乱数を用いてランダムに行う。データベース管理部 80 が母体データ群からデータをランダムに抽出することができない場合、次の手順でデータの抽出を行う。

【0059】

まず、データベース管理部 80 に対して SQL 問い合わせを行い、母体データ群における各属性値組合せ（キー）に合致するデータと抽出量に関する情報を取得する（図 7、ステップ 235）。図 20 は、このようにして取得されたキーデータ及び抽出量の一覧を示す図表である。図示の例では、キー 11111 からキー 99999 までの 9 個のキーと、このキーに該当するデータの抽出量とが対応付けられている。このデータは、乱数順キー記憶部 50 に格納される。

【0060】

次に、データ処理部 20 は、乱数順キー記憶部 50 に格納されている、図 20 に示すキーデータのうちの二つのキーデータをランダムに選択し、位置を入れ替える（ステップ 236）。この操作を一定回数繰り返し、図 20 に示したキーデータの並びをランダムな並びに変換する（ステップ 237）。図 21 は、このようにして作成されたランダムなキーデータ及び抽出量の一覧である。

【0061】

データ処理部 20 は、目標抽出量分のデータを抽出し、抽出したデータ群（結果データ群）を抽出結果データベース 70 に格納する（ステップ 238、239）。データの抽出は、データベース管理部 80 への SQL 問い合わせにより行う。

そして、全ての属性値組合せに対応するデータを、それぞれ目標抽出分だけランダムに抽出し、抽出結果データベース 70 に格納したならば、データ抽出処理を終了する（ステップ 240）。

【0062】

以上のようにして取得された結果データ群は、抽出結果データベース 70 に蓄積される。そして、外部から所定の結果データ群に対するデータ取得要求が入力

された場合に、データ処理部 2 0 が、データベース管理部 8 0 を介して抽出結果データベース 7 0 から該当する結果データ群を読み出し、出力部 9 0 を用いて出力する。

【 0 0 6 3 】

上述したデータベースシステムを債権流動化に伴う流動化商品の選定において、金融機関などの債権の商品化を行おうとしている主体が、保有している債権の集合の中から商品化の対象となる債権の集合を抽出するために用いられれば、当該商品における債権の属性の比率を任意に指定することができ、容易に当該比率で債権が配分された商品を作成することができる。

【 0 0 6 4 】

【発明の効果】

以上説明したように、本発明によれば、2 次元以上の特定の構成比を持つ結果データを選択する作為的抽出を正しく行うことが可能となる。

【図面の簡単な説明】

【図 1】 本実施の形態におけるデータベースシステムの全体構成を説明する図である。

【図 2】 本実施の形態によるデータ抽出の動作を説明するフローチャートであり、構成条件の属性値ごとの目標抽出量と抽出可能量との対応データを作成する動作を説明する図である。

【図 3】 本実施の形態によるデータ抽出の動作を説明するフローチャートであり、属性値組合せごとの目標抽出量と抽出可能量との対応データを作成する動作を説明する図である。

【図 4】 本実施の形態によるデータ抽出の動作を説明するフローチャートであり、対角入替調整の対象を決定する動作を説明する図である。

【図 5】 本実施の形態によるデータ抽出の動作を説明するフローチャートであり、対角入替調整の動作を説明する図である。

【図 6】 本実施の形態によるデータ抽出の動作を説明するフローチャートであり、対角入替調整を全ての属性値組合せのデータについて再帰的に行うための動作を説明する図である。

【図 7】 本実施の形態によるデータ抽出の動作を説明するフローチャートであり、母体データ群から実際にデータを抽出する動作を説明する図である。

【図 8】 本実施の形態における抽出条件入力部より入力される、全体抽出量、抽出条件、構成比の一例を示す図表である。

【図 9】 構成条件及び属性値ごとの目標抽出量と抽出可能量との対応データの一例を示す図表である。

【図 1 0】 図 9 の対応データにおいて全体抽出量を調整した状態を示す図表である。

【図 1 1】 図 1 0 の対応データに基づいて算出された属性値組合せごとのバランス比率と目標抽出量と抽出可能量との対応データを示す図表である。

【図 1 2】 本実施の形態における対角入替調整の基本概念を説明する図表である。

【図 1 3】 図 1 1 に示した対応データを、各構成条件を座標軸とする 3 次元空間として認識したイメージを示す図表である。

【図 1 4】 図 1 3 の対応データに対して対角入替調整を行った様子を示す図表である。

【図 1 5】 図 1 4 の対応データに対して対角入替調整をさらにを行った様子を示す図表である。

【図 1 6】 図 1 5 の対応データに対して対角入替調整をさらにを行った様子を示す図表である。

【図 1 7】 図 1 1 に示した属性値組合せごとの対応データに、図 1 6 までの対角入替調整を行った後の目標抽出量を加えた対応データを示す図表である。

【図 1 8】 図 1 6 の対応データに対して対角入替調整をさらにを行った様子を示す図表である。

【図 1 9】 図 1 1 に示した属性値組合せごとの対応データに、図 1 8 までの対角入替調整を行った後の目標抽出量を加えた対応データを示す図表である。

【図 2 0】 本実施の形態において母体データ群から取得されたキーデータ及び抽出量の一覧を示す図表である。

【図 2 1】 図 2 0 のキーデータの一覧をランダムに並べ替えた状態を示す

図表である。

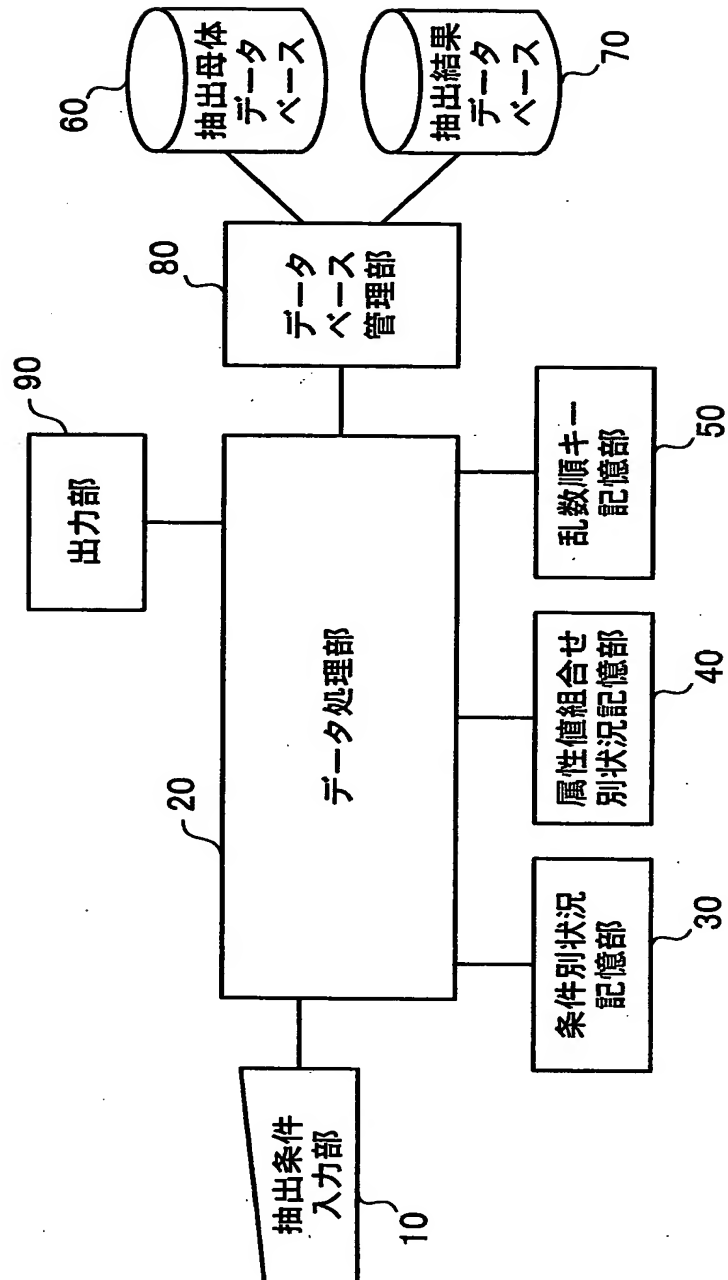
【符号の説明】

1 0 …抽出条件入力部、2 0 …データ処理部、3 0 …条件別状況記憶部、4 0 …属性値組合せ別状況記憶部、5 0 …乱数順キー記憶部、6 0 …抽出母体データベース、7 0 …抽出結果データベース、8 0 …データベース管理部、9 0 …出力部

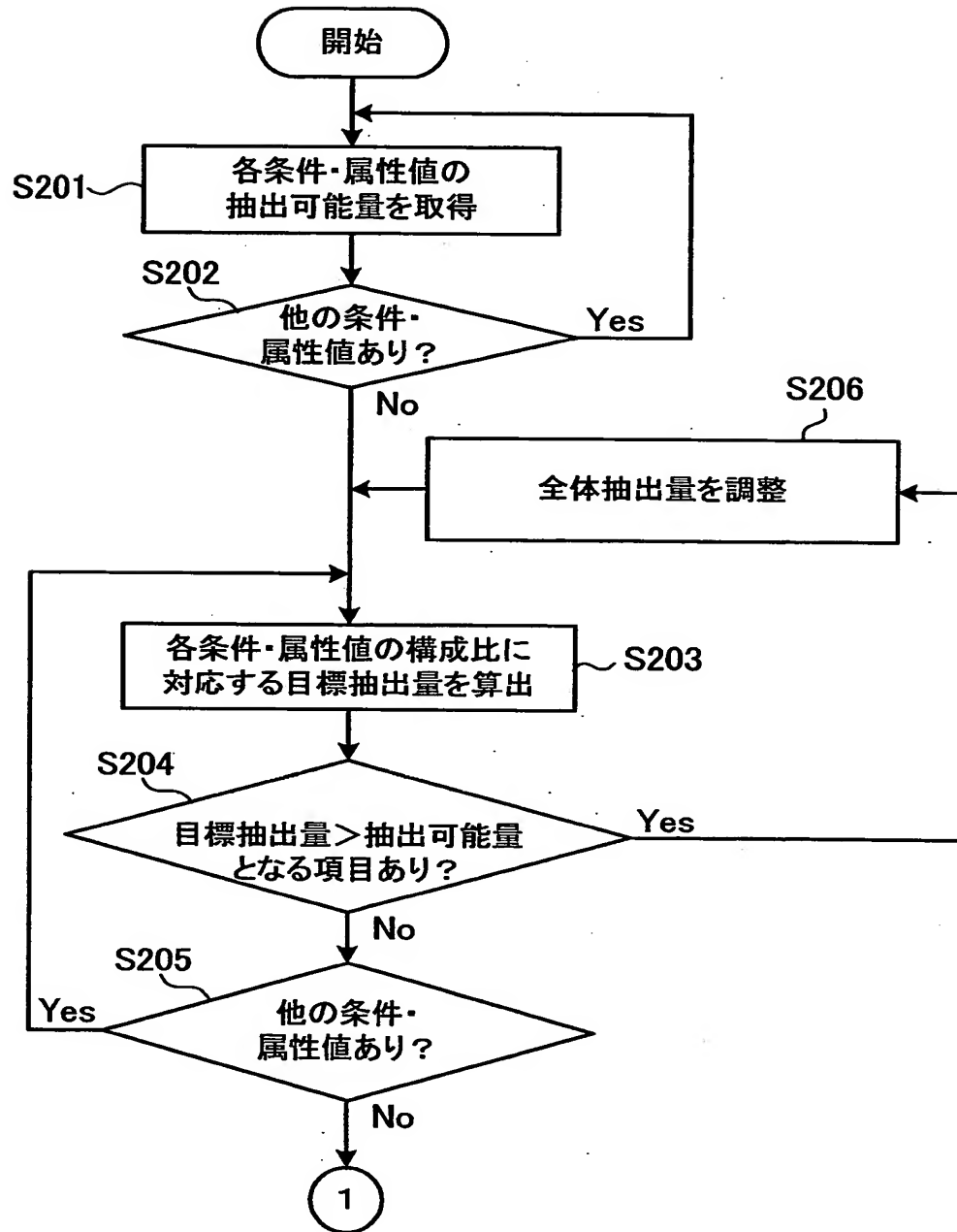
【書類名】

図面

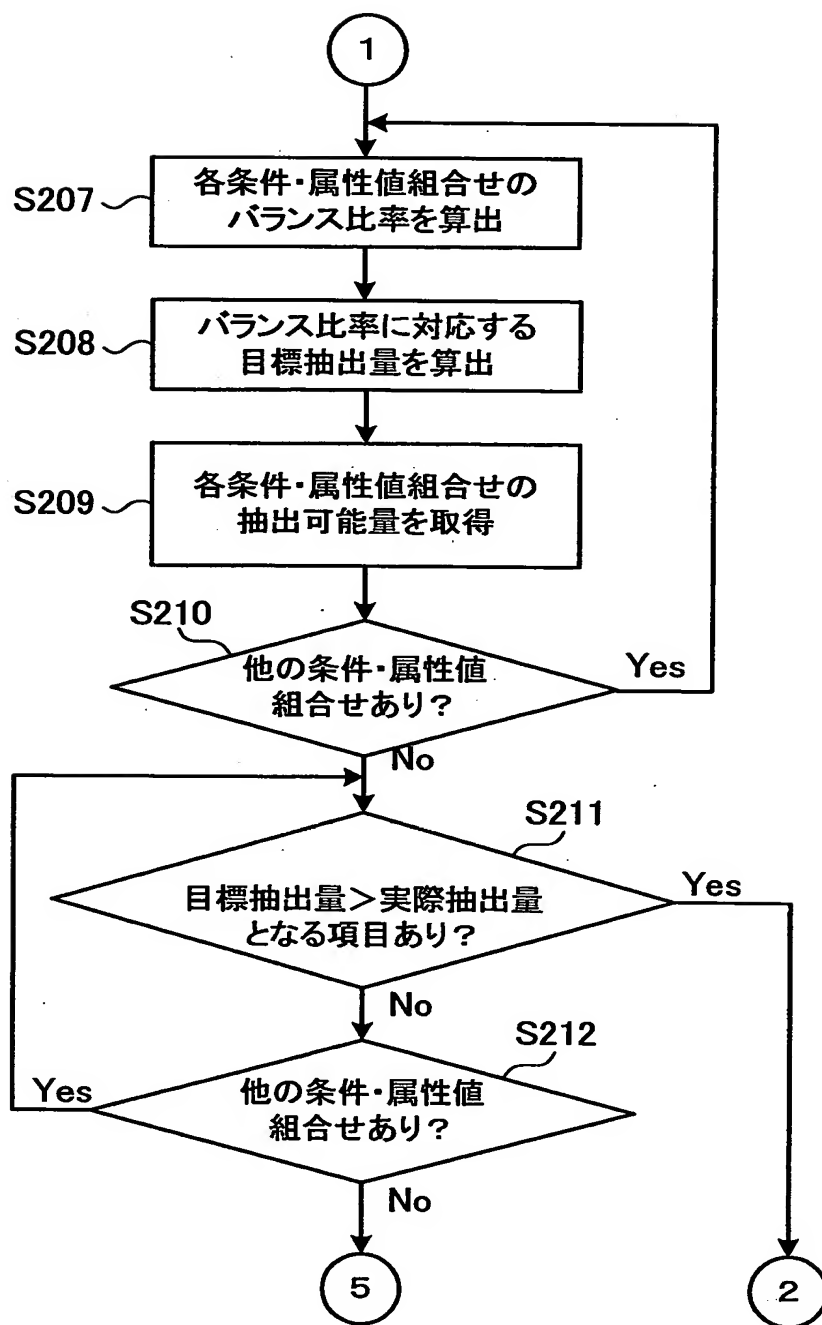
【図 1】



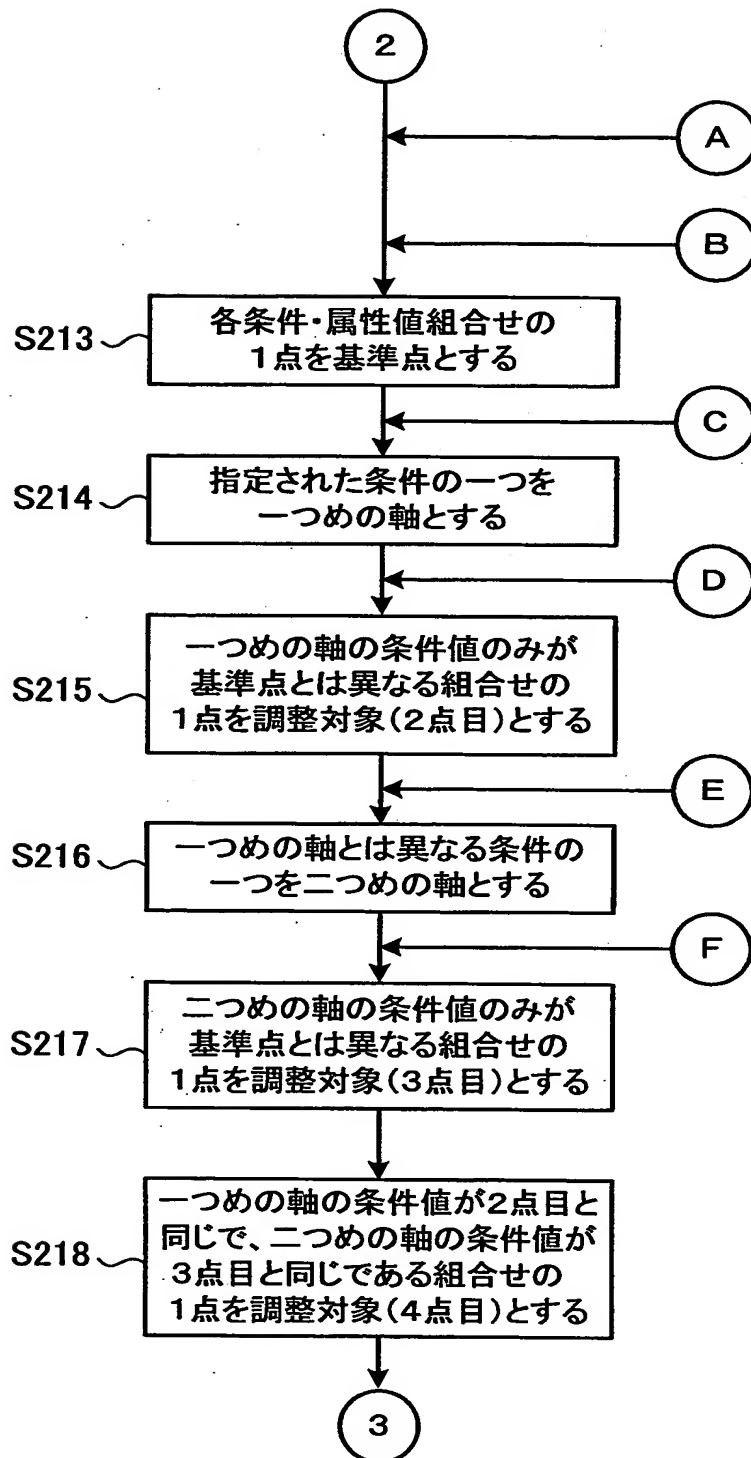
【図 2】



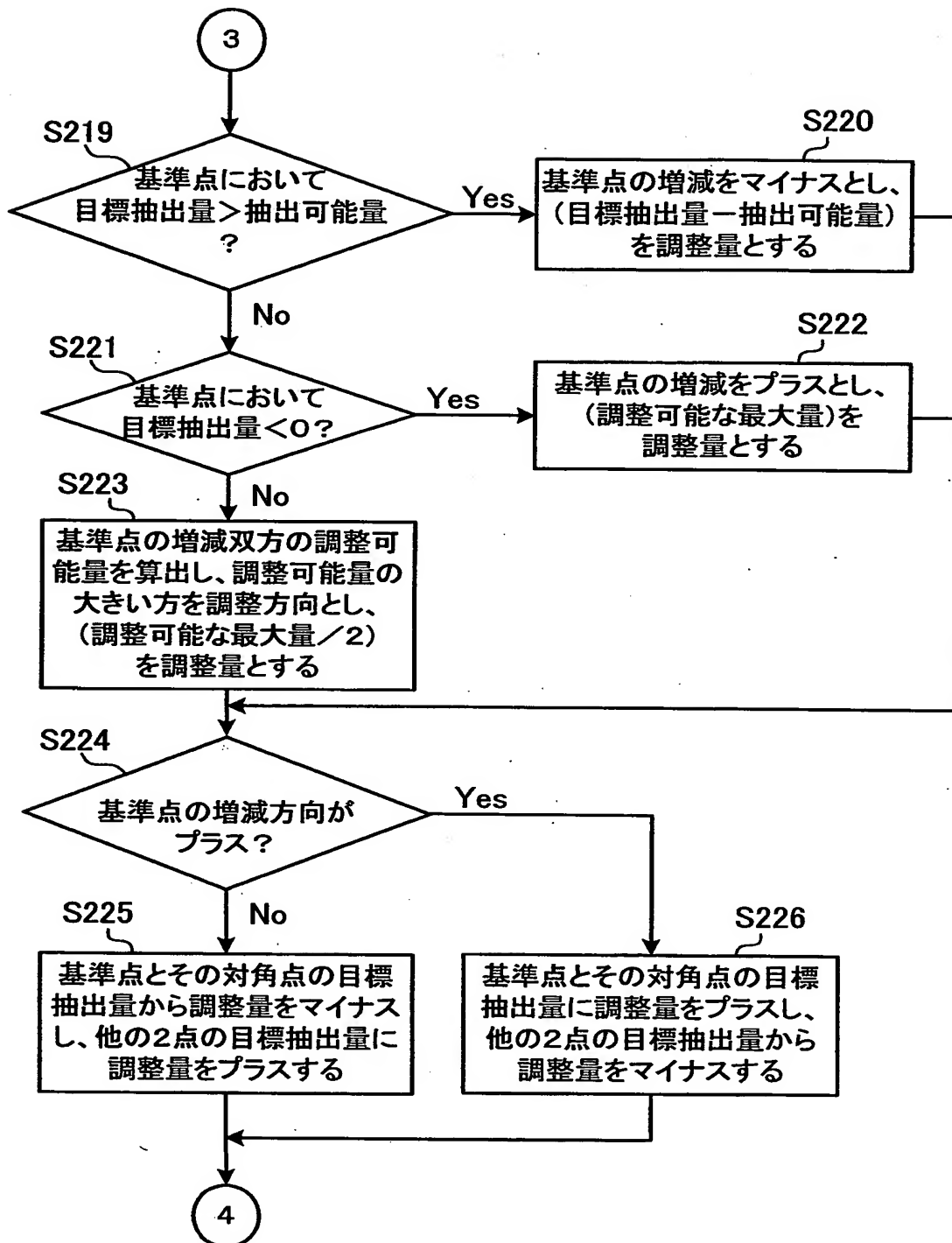
【図 3】



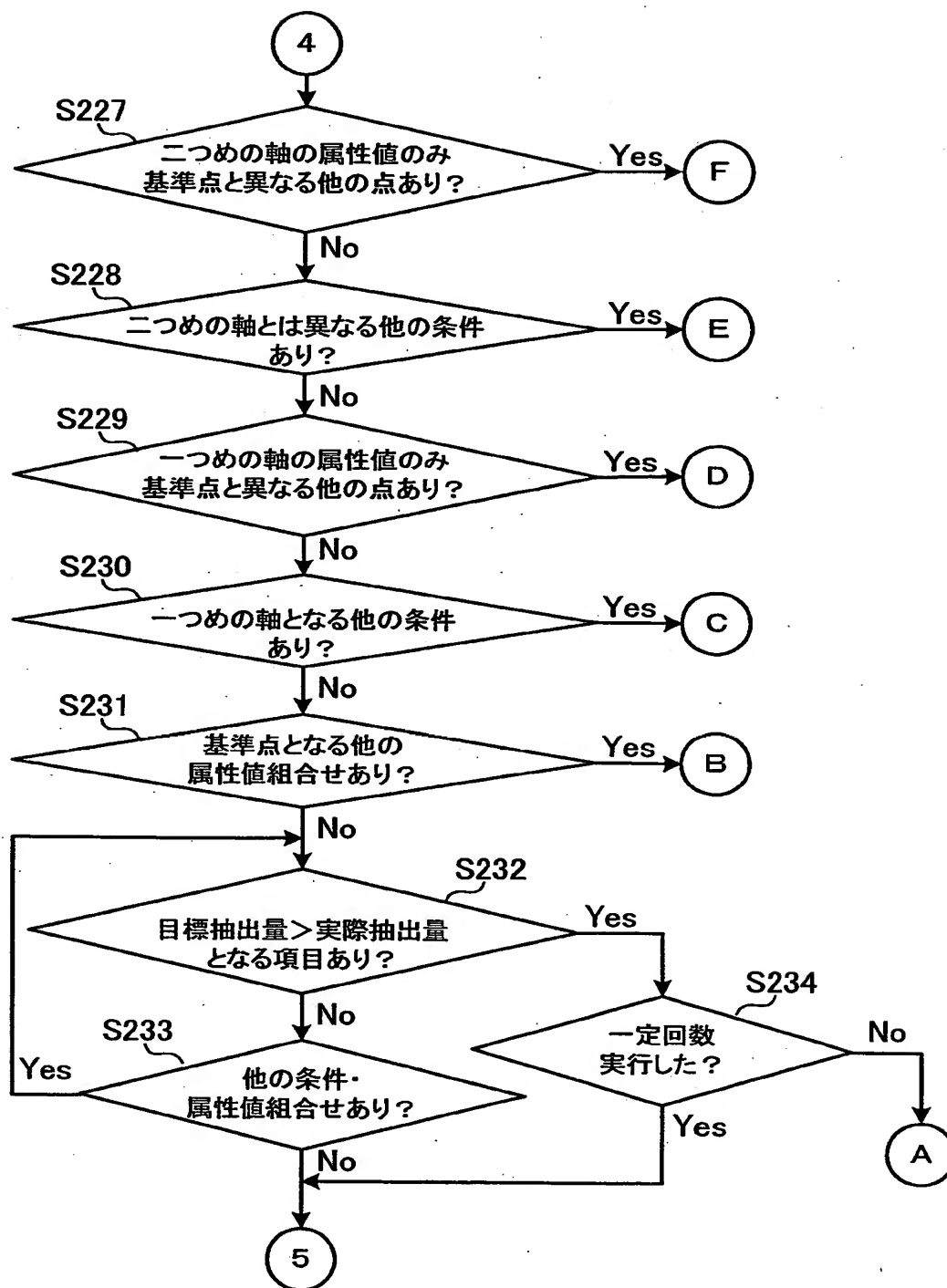
【図4】



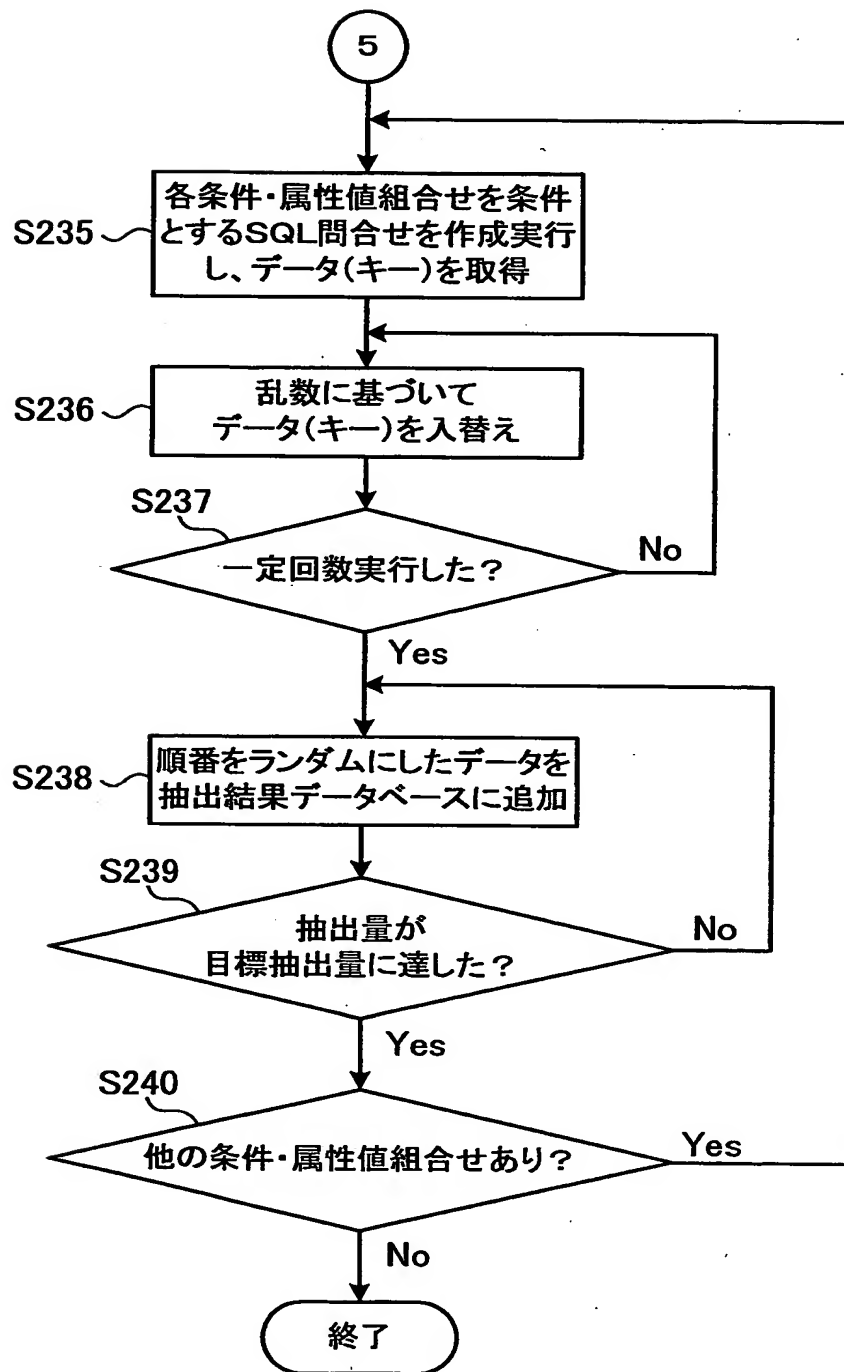
【図 5】



【図 6】



【図 7】



【図 8】

全体抽出量		1 0 0 0									
		A		構成条件 フィールド		B		構成条件 フィールド		C	
構成条件 フィールド		構成比		属性値		構成比		属性値		構成比	
A 1		4 0 %		B 1		2 0 %		C 1		5 0 %	
A 2		6 0 %		B 2		3 0 %		C 2		5 0 %	
				B 3		5 0 %					

【図 9】

全体抽出量	1 0 0 0	
構成条件 フィールド	A	
属性値	目標抽出量	抽出可能量
A 1	4 0 0	5 0 0 0
A 2	6 0 0	5 0 0 0
構成条件 フィールド	B	
属性値	目標抽出量	抽出可能量
B 1	2 0 0	6 0 0
B 2	3 0 0	9 0 0 0
B 3	5 0 0	4 0 0
構成条件 フィールド	C	
属性値	目標抽出量	抽出可能量
C 1	5 0 0	9 7 0 0
C 2	5 0 0	3 0 0

【図 1 0】

全体抽出量		6 0 0
構成条件 フィールド	A	
属性値	目標抽出量	抽出可能量
A 1	2 4 0	5 0 0 0
A 2	3 6 0	5 0 0 0
構成条件 フィールド	B	
属性値	目標抽出量	抽出可能量
B 1	1 2 0	6 0 0
B 2	1 8 0	9 0 0 0
B 3	3 0 0	4 0 0
構成条件 フィールド	C	
属性値	目標抽出量	抽出可能量
C 1	3 0 0	9 7 0 0
C 2	3 0 0	3 0 0

【図 11】

全体抽出量			600		
属性値組合せ			バランス 比率	目標抽出量	抽出可能量
A	B	C			
A1	B1	C1	4%	24	500
A1	B1	C2	4%	24	0
A1	B2	C1	6%	36	4500
A1	B2	C2	6%	36	0
A1	B3	C1	10%	60	0
A1	B3	C2	10%	60	0
A2	B1	C1	6%	36	100
A2	B1	C2	6%	36	0
A2	B2	C1	9%	54	4500
A2	B2	C2	9%	54	0
A2	B3	C1	15%	90	100
A2	B3	C2	15%	90	300

【図12】

(A)

$\alpha \backslash \beta$	β			
		$\beta 1$	$\beta 2$	計
$\alpha 1$		16	24	40
$\alpha 2$		12	18	30
$\alpha 3$		12	18	30
計		40	60	100

Diagram (A) shows adjustments to the values in the table above. Arrows indicate the following changes:

- From $\beta 1$ to $\alpha 1$: +2
- From $\beta 2$ to $\alpha 1$: -2
- From $\beta 1$ to $\alpha 3$: -2
- From $\beta 2$ to $\alpha 3$: +2

(B)

$\alpha \backslash \beta$	β			
		$\beta 1$	$\beta 2$	計
$\alpha 1$		18	22	40
$\alpha 2$		10	20	30
$\alpha 3$		12	18	30
計		40	60	100

【図 1 3】

A 1	C 1		C 2	
	目標抽出量	抽出可能量	目標抽出量	抽出可能量
B 1	2 4	5 0 0	2 4	0
B 2	3 6	4 5 0 0	3 6	0
B 3	6 0	0	6 0	0
A 2	C 1		C 2	
	目標抽出量	抽出可能量	目標抽出量	抽出可能量
B 1	3 6	1 0 0	3 6	0
B 2	5 4	4 5 0 0	5 4	0
B 3	9 0	1 0 0	9 0	3 0 0

【図 1 4】

A 1	C 1		C 2	
	目標抽出量	抽出可能量	目標抽出量	抽出可能量
B 1	2 4	5 0 0	2 4	0
B 2	3 6	4 5 0 0	3 6	0
B 3	6 0	0	6 0	0
A 2	C 1		C 2	
	目標抽出量	抽出可能量	目標抽出量	抽出可能量
B 1	7 2	1 0 0	0	0
B 2	5 4	4 5 0 0	5 4	0
B 3	5 4	1 0 0	1 2 6	3 0 0

【図 1 5】

A 1	C 1		C 2	
	目標抽出量	抽出可能量	目標抽出量	抽出可能量
B 1	2 4	5 0 0	2 4	0
B 2	3 6	4 5 0 0	3 6	0
B 3	6 0	0	6 0	0
A 2	C 1		C 2	
	目標抽出量	抽出可能量	目標抽出量	抽出可能量
B 1	7 2	1 0 0	0	0
B 2	1 0 8	4 5 0 0	0	0
B 3	0	1 0 0	1 8 0	3 0 0

【図 1 6】

A 1	C 1		C 2	
	目標抽出量	抽出可能量	目標抽出量	抽出可能量
B 1	2 4	5 0 0	2 4	0
B 2	9 6	4 5 0 0	3 6	0
B 3	0	0	6 0	0
A 2	C 1		C 2	
	目標抽出量	抽出可能量	目標抽出量	抽出可能量
B 1	7 2	1 0 0	0	0
B 2	4 8	4 5 0 0	0	0
B 3	6 0	1 0 0	1 8 0	3 0 0

【図 17】

全体抽出量			600			
属性値組合せ			バランス 比率	目標抽出量	抽出可能量	調整後 目標抽出量
A	B	C				
A1	B1	C1	4%	24	500	24
A1	B1	C2	4%	24	0	24
A1	B2	C1	6%	36	4500	96
A1	B2	C2	6%	36	0	36
A1	B3	C1	10%	60	0	0
A1	B3	C2	10%	60	0	60
A2	B1	C1	6%	36	100	72
A2	B1	C2	6%	36	0	0
A2	B2	C1	9%	54	4500	108
A2	B2	C2	9%	54	0	0
A2	B3	C1	15%	90	100	0
A2	B3	C2	15%	90	300	180

【図 1 8】

A 1	C 1		C 2	
	目標抽出量	抽出可能量	目標抽出量	抽出可能量
B 1	6 0	5 0 0	0	0
B 2	1 8 0	4 5 0 0	0	0
B 3	0	0	0	0
A 2	C 1		C 2	
	目標抽出量	抽出可能量	目標抽出量	抽出可能量
B 1	6 0	1 0 0	0	0
B 2	0	4 5 0 0	0	0
B 3	0	1 0 0	3 0 0	3 0 0

【図 19】

全体抽出量			600			
属性値組合せ			バランス 比率	目標抽出量	抽出可能量	調整後 目標抽出量
A	B	C				
A1	B1	C1	4%	24	500	60
A1	B1	C2	4%	24	0	0
A1	B2	C1	6%	36	4500	180
A1	B2	C2	6%	36	0	0
A1	B3	C1	10%	60	0	0
A1	B3	C2	10%	60	0	0
A2	B1	C1	6%	36	100	60
A2	B1	C2	6%	36	0	0
A2	B2	C1	9%	54	4500	0
A2	B2	C2	9%	54	0	0
A2	B3	C1	15%	90	100	0
A2	B3	C2	15%	90	300	300

【図 2 0】

キ一	抽出量
1 1 1 1 1	1 2
2 2 2 2 2	5
3 3 3 3 3	1 0
4 4 4 4 4	3
5 5 5 5 5	6
6 6 6 6 6	2
7 7 7 7 7	4
8 8 8 8 8	1 8
9 9 9 9 9	5

【図 2 1】

キ一	抽出量
7 7 7 7 7	4
4 4 4 4 4	3
9 9 9 9 9	5
1 1 1 1 1	1 2
3 3 3 3 3	1 0
2 2 2 2 2	5
6 6 6 6 6	2
8 8 8 8 8	1 8
5 5 5 5 5	6

【書類名】 要約書

【要約】

【課題】 データベースから一定量のデータを抽出する手法として、2次元以上の特定の構成比を持つ結果データを選択する作為的抽出を正しく行うことができるデータ抽出方法を提供する。

【解決手段】 母体データ群から所定のデータの集合を抽出するデータ抽出方法において、抽出すべきデータ集合の要素であるデータの属性の構成比を設定するステップと、母体データ群中に存在するこの属性を持つ対応データの量である抽出可能量を取得するステップと、設定された構成比に基づいてこの属性ごとに母体データ群から抽出すべきデータ量である抽出データ量を計算するステップと、この構成比に基づいて算出された抽出データ量のうち、所定の属性における抽出データ量がこの属性を持つ対応データの抽出可能量を上回っている場合に、この抽出データ量がこの抽出可能量以下となり、かつこの構成比を一定の範囲で保持するように、この抽出データ量を調整するステップとを含む。

【選択図】 なし

特2000-380199

出 願 人 履 歴 情 報

識別番号 [592073101]

1. 変更年月日	1992年 4月 3日
[変更理由]	新規登録
住 所	東京都港区六本木3丁目2番12号
氏 名	日本アイ・ビー・エム株式会社

出 願 人 履 歴 情 報

識別番号 [393000054]

1. 変更年月日 1992年12月10日

[変更理由] 新規登録

住 所 大阪府大阪市中央区備後町2丁目2番1号
氏 名 株式会社大和銀行